Chapter 4

ORGANIZATION OF RETROTRANSPOSONS AND MICROSATELLITES IN CEREAL GENOMES

Alan H. Schulman[1,*], Pushpendra K. Gupta[2] and Rajeev K. Varshney[3]

[1]*MTT Agrifood Research, Jokioinen, and University of Helsinki, Helsinki, Finland*; [2]*Ch. Charan Singh University, Meerut, U.P., India*; and [3]*Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany*
[*]Author for correspondence: alan.schulman@helsinki.fi

1. INTRODUCTION

Our understanding of genome organization has its roots in postwar interest in the effects of radiation. Observations on the relationship between doses of ionizing radiation and the frequency of mutations (Abrahamson *et al.*, 1973; Trujillo and Dugan, 1975) indicated that the size of the genetic target receiving the radiation dose varied considerably between organisms. By the beginning of the 1970s, this phenomenon had come to be known as the "*C*-value paradox" (Thomas, 1971). The paradox was that the total genome size, or *C*-value, varied widely within a given clade of organisms and bore no relationship to organismal complexity. For example, two legumes within the same genus, *Vicia faba* and *Vicia sativa*, have haploid genomes of 13.1 x 10^9 and 2.2 x10^9 respectively, but differ very little morphologically. This observation has been fully confirmed by the large-scale determination of genome sizes of many plants (http://www.rbgkew.org.uk/cvalues), within which genome size varies from about 10^7 bp in *Cardamine* and Arabidopsis among the Crucifereae to nearly 10^{11} bp in *Fritillaria* among the Lilliaceae. Within the cereals, rice (*Oryza sativa*) has a compact genome of 4.8 x 10^8 bp, ranging upwards through sorghum (7.35 x 10^8), maize (26.7 x 10^8), and barley (54.4 x 10^8 bp).

P K. Gupta and R K. Varshney (eds), Cereal Genomics, 83-118.

Resolution of the *C*-value paradox began with experiments carried out in the 1970s and early 1980s (*e.g.*, Flavell *et al.*, 1974; Hake and Walbot, 1980). Sheared DNA was melted and then allowed to reanneal over protracted periods of time to various C_0t (combinations of concentration and time) levels. The experiments showed that genomes of higher organisms are comprised of single-copy DNA regions interspersed between repetitive DNA of various degrees of redundancy. These and later studies have led to the view that, in general, larger eukaryotic genomes have a high proportion of repetitive DNA. The maize nuclear genome for instance, contains about 60 to 80% repetitive DNA (Flavell *et al.*, 1974; Bennetzen *et al.*, 1994; Springer *et al.*, 1994; Heslop-Harrison, 2000; Meyers *et al.*, 2001) and the wheat nuclear genome contains 83 % repetitive DNA (Wicker *et al.*, 2001). These studies were, however, essentially static views of the genome lacking details of structure and function, and left unexplained why, within narrow clades of plants, genome sizes should show such variation.

A final resolution of the question of why eukaryotic genomes might vary in size derives from several fields of research. First, large-scale sequencing of ESTs from many plants and of the entire genomes of mouse, human, Arabidopsis, and rice indicates that the number of cellular genes is relatively constant, some 30,000 to 50,000 (TAGI, 2000; Goff *et al.*, 2002; Kurata *et al.*, 2002; Yu *et al.*, 2002). Given 30,000 genes with 4800 bp, 60 % of the rice genome but only 2.7 % of the barley genome encodes cellular functions.

Second, sequencing of long contiguous segments ("contigs") of the genome, such as in assembled BAC and YAC clones (respectively, Bacterial Artificial Chromosome and Yeast Artificial Chromosome), confirmed the hybridizaton results that much of the intervening DNA between cellular genes consists of repetitive DNA. Repetitive sequences can be found in the genome either in tandem arrays or in a dispersed fashion and can be classified into 3 categories: (a) transposable elements, which are mobile genetic elements, largely consisting of retrotransposons, in many plant species; (b) microsatellite sequences, which are tandemly repeated DNA sequences (also called simple sequence repeats); and (c) a special class of repetitive sequences including telomeric and centromeric sequences as well as tandem arrays of rDNA units.

Microsatellites and particularly retrotransposons are dynamic components of the genome whose abundance can change over time. As detailed below, the dynamics of retrotransposon replication and insertion between cellular genes can explain much of the size differences between otherwise similar plant genomes. Furthermore, cereal genomes are largely syntenic and colinear, having largely the same genes on homeologous chromosomes in mostly the

same order (Smilde *et al.*, 2001). The major part of cereal genomes is comprised of retrotransposons and microsatellites, and the major factor in the differences between their genome sizes is the varying proportions of these elements. Therefore, in the present chapter, we discuss the distribution and organization of retrotransposons and microsatellites and their impact on genome size in cereals.

2. RETROTRANSPOSONS AND THEIR ORGANIZATION WITHIN THE GENOME

Although the study of retroelements can be dated to about 1910 with the discoveries of avian leucosis virus (alv) and Rous sarcoma virus (Rous, 1911), their true nature was mysterious before the provirus hypothesis of Temin (Temin, 1964) was confirmed with the discovery of reverse transcriptase (Baltimore, 1970; Temin and Mizutani, 1970). The mode of retroviral replication was worked out during the 1970s, culminating in the first sequence of a retrovirus (Shinnick *et al.*, 1981). The presence of retroelements in vertebrate genomes, termed endogenous retroviruses, was subsequently established (Löwer *et al.*, 1996). By 1985, yeast was demonstrated to have endogenous elements, which were named *Ty*, transposing, as do the retroviruses, through an RNA intermediate (Boeke *et al.*, 1985; Kingsman and Kingsman, 1988). The presence and activity of similar elements in insects (Flavell *et al.*, 1980; Levis *et al.*, 1980; Shiba and Saigo, 1983) and in the plants (Shepherd *et al.*, 1984; Grandbastien *et al.*, 1989) was shown in the same period. The organization of retrotransposons in plant genomes has been investigated on four levels of increasing precision: Southern- and dot-blot hybridization studies giving a general view of copy number; *in situ* hybridization of chromosomes, yielding a picture of the distribution by chromosome region; mapping and sequencing of large contiguous segments of the genome, such as are present in BACs and YACs, detailing local organization; large-scale sequencing extending over whole genomes, making possible both a detailed local and global view.

2.1. Retrotransposon Structure

Retrotransposons are transposable elements that replicate via an RNA intermediate in a process involving transcription and reverse transcription, and are labeled Class I. Class II transposable elements are DNA transposons, which replicate only as part of the chromosome and move by a cut-and-paste mechanism. The Class I elements fall into three distinct groups. These are:

LINEs, or long-interspersed elements; SINEs, or short interspersed elements; LTR retrotransposons, which are bounded by long terminal repeats (LTRs). The LTR retrotransposons can be further divided into two categories, the *copia*-like and the *gypsy*-like elements. Members of each class display a greater similarity to members of the same class from different species and Kingdoms than to elements from the same species, supporting their ancient origin (Xiong and Eickbush, 1990). The three retrotransposon groups found in plants each occur in the animals and fungi as well, and therefore represent ancient genomic components that predate the divergence of these three major groups of eukaryotes.

The structure and life cycle of the plant LTR retrotransposons (reviewed in Kumar and Bennetzen, 1999) most resembles that of the mammalian retroviruses. The bounding LTRs contain the signals needed for RNA expression and processing. The 5' LTR serves as the promoter, whereas the 3' LTR functions as the terminator and polyadenylation signal. The LTRs can vary in length from several 100 bp (*Tos17*; Hirochika *et al.*, 1992) to over 5 kb (*Sukkula*; Shirasu *et al.*, 2000) and are themselves terminated by small inverted repeats giving the LTRs a universal 5' TG....CA 3' structure. Adjacent, respectively, to the 5' and 3' LTRs are the priming sites for the (-)-strand and (+)-strands of the cDNA that are recognized by the reverse transcriptase (RT) encoded by the retrotransposon. In addition to RT, LTR retrotransposons encode the same basic proteins, generally excepting the envelope protein, as do retroviruses: capsid protein (GAG), which forms the virus-like particle (VLP), aspartic proteinase (AP), which cleaves the expressed polyprotein into functional components, integrase (IN), which carries out the insertion of the cDNA into the genome, and RNase H (RH), which is required in replication (Turner and Summers, 1999). In the *copia*-like retrotransposons, the domains are organized 5' LTR-GAG-AP-IN-RT-RH-LTR 3', whereas *gypsy*-like elements are arranged as 5' LTR-GAG-AP-RT-RH-IN-LTR 3'. The GAG is often, though not necessarily, in a different reading frame than the other domains, which are expressed as single polyprotein referred to as POL. A newly recognized subset of the plant *gypsy*-like retrotransposons contains a coding domain (*env*), internal to the 3' LTR, specifying a putative envelope protein (Vicient, *et al.*, 2001a). Similar domains have been found in *copia*-like elements as well (Laten *et al.*, 1998; Wright and Voytas, 2002). The *env*-bearing plant elements of the *gypsy*-like class are organized in a manner identical to the mammalian retroviruses, raising the still-unanswered question of their function as viruses. A new class of non-autonomous elements derived from retrotransposons but lacking protein-coding domains has been reported (Witte *et al.*, 2001).

The LTR retrotransposons are the most abundant and best-studied class of plant retroelements, but the LINEs and SINEs are important genomic components as well (Noma *et* al., 1999; Schmidt, 1999; Vershinin *et al.*, 2002). The LINE elements are viewed as progenitors of the LTR retrotransposons, although the two groups replicate by different mechanisms (Eickbush, 1992). The LINE elements lack both LTRs and IN, and at least in mammals replicate via transcription from an internal promoter, followed by reverse transcription primed by genomic DNA at a cleaved target site (Luan *et al.*, 1993; Weiner, 2002). Integration likely takes place via a break repair mechanism (Moore and Haber, 1996). In humans, the L1 family of LINE elements is a major mutagenic agent and responsive for genome remodeling via extensive deletions, at least in transformed cells (Kazazian and Goodier, 2002). Although few plant LINEs appear functional (Schmidt, 1999), the recently reported Karma element in rice can be activated by induced hypomethylation in tissue culture, leading to copy number increases in subsequent generations (Komatsu *et al.*, 2003).

The SINEs, in contrast to LINEs, encode no enzymatic functions, and appear to rely on those of LINEs for their propagation (Boeke, 1997; Lenoir *et al.*, 2001). The SINEs are reverse-transcribed from RNA polymerase III products and, with the exception of some mammalian elements, are derived from tRNA; as such, they contain internal promoters that can render them transcriptionally active elements (Deragon and Capy, 2000). The SINE elements are mobile within the plants, and have been well studied in the genus Brassica (Gilbert *et* al., 1997; Arnaud *et al.* 2000; Tikhonov *et al.*, 2001). They have been used for phylogenetic markers in both Brassica and *Oryza*, in humans, and in many animals (Cheng *et al.*, 2002).

Since the discovery of the first active plant retrotransposon (Grandbastien *et al.*, 1989), considerable progress has been made in demonstrating that the various steps of the life cycle of LTR retrotransposons previously established for yeast and flies are carried out in the plants (Kumar and Bennetzen, 1999; Vicient *et al.*, 1999a). Among the cereals, transcription of retrotransposons appears to be a common phenomenon, corresponding to about 0.1% of total transcripts in EST databases (Vicient *et al.,* 2001b; Echenique, *et al.*, 2002). In the cereals, transcription has been directly demonstrated in rice (Hirochika, 1993; Hirochika *et al.,* 1996), maize (Turcich *et al.*, 1996), and barley (Suoniemi *et al.,* 1996a; Vicient *et al.,* 2001a). Translation, polyprotein processing, and formation of virus-like particles have been explicitly demonstrated in barley and other cereals (Jääskeläinen *et al.,* 1999; Vicient *et al.* 2001b) for the abundant *BARE*-1 retrotransposon family (Manninen and Schulman, 1993; Vicient *et al.*, 1999a) and transcriptional splicing has been for the retroviral-like *Bagy2* element (Vicient *et al.*, 2001a). Stress and tissue

culture activation of retrotransposons appears to be a general phenomenon (Wessler, 1996; Grandbastien, 1998), and has been well analyzed in tobacco and rice (Hirochika *et al.*, 1992; Hirochika *et al.*, 1996; Takeda *et al.*, 1998; Beguiristain *et al.*, 2001). In barley, although stress activation has not been directly demonstrated, both retrotransposon *BARE*-1 copy number and genome size is correlated with environmental factors associated with drought and temperature stress (Kalendar *et al.*, 2000).

2.2. Retrotransposon Copy Number Variation and its Relationship to Genome Size

Given actively replicating families of retrotransposons in plant genomes and gene numbers varying over a relatively small range, much of the considerably greater variation observed in genome size among the flowering plants can be attributed to variation in the bulk contribution of retrotransposons to the total genome. This appears true for eukaryotes generally; the compact genome of the yeast *Saccharomyces cerevisiae*, comprising 13 x 10^6 bp, contains 331 retroelements, of which only 51 are full-length retrotransposons (Kim *et al.*, 1998). The great majority of the elements in yeast are, instead, solo LTRs, the remnant of recombinational loss of a retrotransposon. Occurrence of solo LTRs shall be revisited below with respect to the cereals. Taken together, the full-length and solo LTRs comprise 3.1 % of the yeast genome.

The genome of *Arabidopsis thaliana*, at 10^8 bp, is the smallest described in the plants (Goodman *et al.*, 1995). A careful survey of 1.7 x 10^7 bp (Le *et al.*, 2000) revealed the presence of 50 groups of LTR retrotransposons, divided about equally between gypsy-like and copia-like elements, each with about two members on average. In addition, three groups, or more correctly two groups (Lenoir *et al.*, 2001), of SINEs and 28 groups of LINEs were found, the latter with about one member per group. Extrapolating from the 134 Class I elements identified to the whole genome yields an estimated total of 779 elements in *A. thaliana*, comprising about 2 % of the DNA.

The cereals tend to have large genomes, and commensurately more retrotransposons. Cultivated rice has one of the smallest genomes among the cereals, containing 4.3 x 10^8 bp (Kurata *et al.*, 1997). Dispersed within the rice genome is about 10^3 retrotransposons when estimated by hybridization with a conserved probe (Hirochika *et al.*, 1992). The rice genome project has recently completed three chromosomes representing 100.18 Mb to date, about 23% of the total (Sasaki *et al.*, 2002; Feng *et al.*, 2002; The Rice

Chromosome 10 Sequencing Consortium, 2003). Annotation of the amassed sequence (supplementary material at http://sciencemag.org/cgi/data/300/5625/1566/DC1/1) reveals 11,423 retrotransposons covering 8.27 Mb. Extrapolating to the whole genome, this indicates a total of 4.9 x 10^4 retrotransposons covering 35 Mb, or 8.3% of the entire genome. The discrepancy with earlier estimates may be due in part to the annotation being based on scoring of LTRs rather than coding domains, and on the limitations of hybridization for detecting highly degenerate retroelements. An analysis of the genome size and retrotransposon content of barley, wild barley, and other species within the genus *Hordeum* confirmed the broad correlation of genome size with retrotransposon copy number (Vicient *et al.*, 1999b).

2.3. Chromosomal Distribution of Retrotransposons

The high copy number of retrotransposons in cereal and other genomes raises the question of how their bulk is distributed within the genome. The method of fluorescent *in situ* hybridization (FISH) offers an efficient strategy for acquiring a broad overview of genome organization, and has been applied to many retrotransposon systems. The 1 – 2 x 10^4 copies of the *BARE*-1 retrotransposon, for example, are dispersed along all seven haploid chromosomes of barley and related species, except in the centromeres, telomeres, and nucleolar organizing regions (Suoniemi *et al.*, 1996b; Vicient *et al.*, 1999b). The distribution of the wheat retrotransposon WIS 2-1A, which is sufficiently similar to *BARE*-1 to consider it a member of the *BARE*-1 family (Manninen and Schulman, 1993), was examined in the chromosomes of wheat and rye that are present in the allopolyploid triticale (x*Triticosecale*) by FISH (Muñiz *et al.*, 2001). The elements were dispersed on all chromosomes of both constituent genomes in a manner similar to *BARE*-1 in barley. Likewise, a retroelement fragment from the oat genus *Avena*, 67 % and 70 % similar to *BARE*-1 and WIS-2-1A LTRs respectively, is dispersed along all chromosomes of diploid, tetraploid, and hexaploid species except in the centromeric and nucleolar organizer regions (Katsiotis *et al.*, 1996; Linares *et al.*, 1999). Other *copia*-like (Moore *et al.*, 1991; Svitashev *et al.*, 1994) and *gypsy*-like (Vershinin *et al.*, 2002) elements show a distribution in barley broadly similar to that of *BARE-1*.

Few studies have been carried out on the distribution of the non-LTR retrotransposons, the LINEs and SINEs, in cereal genomes. A comparison of the distribution of *copia*-like, *gypsy*-like, and LINE elements in the wild barley *Hordeum spontaneum*, and in *Aegilops* speltoides, another member of the tribe Triticeae was made (Belyayev *et al.*, 2001). A clustered distribution

of *copia*-like elements was observed in *Ae. speltoides* and two main clusters on different chromosomes found in *H. spontaneum*. The *gypsy*-like elements were present as clusters in both species, whereas LINE elements showed a broader, but nevertheless clustered, distribution. Vershinin and coworkers (2002) reported on LINE distribution in several *Hordeum* species. They observed dispersion along most chromosomes, limited by the low copy numbers in cultivated barley, *H. vulgare*.

Although most families of retrotransposons in the cereals are excluded from the centromere and the nucleolar organizing regions, or are at least rare enough there to give weak FISH signals, some families show converse distributions. One of these, a gypsy-like element named *CEREBA*, is highly localized, but not exclusive, to cereal centromeric regions (Presting *et al.*, 1998; Hudakova *et al.*, 2001). The *gypsy*-like *RIRE7* element of rice also has reported centromeric association (Kumekawa *et al.*, 1999, 2001a; Nomomura *et al.*, 2001). In maize, the CentA retroelement is confined to centromeric regions, whereas the Huck and Prem2 retrotransposons with which it is associated were found not only in the centromeres but also elsewhere on the chromosomes (Ananiev *et al.*, 1998a). In addition, there are a number of other partially characterized retroelements that have been found within cereal centromeres (Dong *et al.* 1998; Miller *et al.* 1998; Nomomura and Kurata, 1999; Langdon *et al.*, 2000; Fukui *et al.*, 2001). A parallel situation occurs in Arabidopsis, where sequenced contigs for the centromeres of chromosomes 4 and 5 reveal a core of 180 bp repeats interrupted by the insertion of *gypsy*-like *Athila* retrotransposons (Kumekawa *et al.*, 2000,2001b). It is tempting to speculate that a single ancient family of *gypsy*-like elements is found in the centromeres of diverse plants, given that such elements can interact directly with conserved kinetochore proteins (Zhong *et al.*, 2002). As the sequences of more centromeres are determined through improved methods and our understanding of centromere function increases (Lamb and Birchler, 2003), a clearer picture will no doubt emerge.

Centromeres, telomeres, and nucleolar organizers have specific functions associated with particular classes of repetitive DNA (Pardue *et al.*, 1996). The absence of all but a few specific families of retrotransposons from these regions may be evidence for strong selective pressure to maintain structural integrity. The telomeres, in particular, appear retrotransposon-poor. However, there are the startling cases of HeT-A and TART in *Drosophila*. These non-LTR retrotransposons are exclusively found at telomeres as long head-to-tail arrays, where they deliver the essential replicative functions needed to maintain the chromosomal ends (Danilevskaya *et al.*, 1998; Casacuberta and Pardue, 2003). Integration at the correct site is orchestrated by the targeting of Gag to the telomeric regions (Rashkova *et al.*, 2003).

Whether similar roles are played by the retrotransposons that are present in structurally critical parts of plant chromosomes, or whether instead those families of retrotransposons are merely not sufficiently deleterious to be excluded, remains to be established.

2.4. Local Organization of Retrotransposons and Genes

Although *in situ* hybridization gives a general view of which compartments in the genome are favored insertion sites for particular classes of retrotransposons, its resolution is too low to reveal detail on how retrotransposons are situated with respect to genes. The advent of large-scale sequencing has now provided an opportunity to carry out in-depth organizational analyses of genes and retrotransposons on contiguous chromosomal segments extending over tens or hundreds of kilobases. The general view has emerged, consistent with earlier data derived from analytical centrifugation, that the genomes of cereals, and grasses more generally, are segmented into compact islands of genes, "gene-space," surrounded by large expanses of repetitive DNA "seas" largely comprised of retrotransposons (Barakat *et al.*, 1997; Sandhu and Gill, 2002; see Chapter 12 of this volume also). These gene blocks appear to vary widely in size. The analysis of a 280 kb segment of the maize genome containing the *Adh1F* allele revealed long stretches of retrotransposons flanking two individual genes and amounting to 60% of the region (SanMiguel *et al.*, 1996). The retrotransposons, furthermore, were present as nests several elements deep, rather than as concatenated blocks. A later analysis of a 225 kb segment in the *Adh* region revealed retrotransposon stretches from 14 to 70 kb surrounding five single genes and a 39 kb block containing four putative genes and no retroelements (Tikhonov *et al.*, 1999). The *bz* locus of maize, which has a very high rate of recombination, is present on a unusually long retrotransposon-free expanse, 32 kb, containing ten genes (Fu *et al.*, 2001).

Detailed analyses of chromosome segments in cereals other than maize have extended and confirmed the model of cereal genomes as consisting of gene islands surrounded by retrotransposon seas. In a 60 kb stretch of the barley genome from chromosome 4HL that contains the *Mlo* locus, the Schulze-Lefert group found three genes clustered in a block of 18 kb (Panstruga *et al.*, 1998). An almost identical arrangement was found on a 66 kb contig from chromosome 2HL, which contains three genes in an interval of 18 kb flanked by a region that has undergone at least 15 retrotransposon integrations (Shirasu *et al.*, 2000). A contig of 261 kb was sequenced from the barley *Mla* locus specifying resistance to powdery mildew (Wei *et al.*, 2002). This segment contains 32 predicted genes organized into three gene islands

separated by two complexes of nested retrotransposons. The *BARE-1* retrotransposon alone accounts for 17.5 % of the total contig. Analysis of four BACs, covering 417.5 kb, confirmed the gene island model of the barley genome (Rostoks *et al.*, 2002). The finishing and full annotation of the rice genome will give the first complete view of the organization of a cereal genome (Goff *et al.*, 2002). Initial glimpses, reconstructions, and annotation of sequenced regions (Nagano *et al.*, 1999; Yu *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortium, 2003) indicate that the organization described for the few contigs of other cereals largely holds for the rice genome as a whole.

Most long segments of DNA that have been analyzed in detail are derived from genic regions. The cloning and sequence assembly of centromeres and other non-genic structures has been generally difficult, and these are under-represented both in general databases and in genome projects. As discussed above, the centromeres of maize contain tandem arrays of CentC elements, interrupted by variable numbers of centromere-specific CentA retrotransposons (Ananiev *et al.*, 1998a). Maize chromosomes possess as well, depending on the variety, unusual heterochromatic elements descriptively called knobs. An analysis of knob sequences revealed that retrotransposons interrupt tandem arrays of diagnostic 180 bp repeats (Ananiev *et al.*, 1998b). Huck elements were not found in any knob segments examined, although other elements such as Prem2 and Zeon were present and Grande seemed particularly abundant.

2.5. Retrotransposons and the Evolution of Genome Organization

Variations in the copy number of retrotransposon families within a species and between related species demonstrate that these elements dynamically contribute to genome evolution over time. To understand these changes, several aspects must be considered: integration site preference; the balance between gain and loss of integrated elements; selective forces acting on both copy number and integration pattern. One approach to reconstructing the evolutionary role of retrotransposons is to compare the sequences of equivalent regions of the genome from different cereals. This is aided by the twin phenomena of colinearity and synteny within the cereals, meaning that the order and content of genes is generally conserved over distances extending to the length of full chromosomes (Moore *et al.*, 1993; Bennetzen and Freeling, 1997). Sorghum and maize are ideal for comparison because of the fine-scale colinearity of their genes despite the maize genome being 3.5-

fold larger (Chen *et al.*, 1997). Sequencing and comparison of the *adh1* region in sorghum and maize showed that the retrotransposon nests comprising over 70% of the contig in maize are all missing from the orthologous region in sorghum (SanMiguel and Bennetzen, 1998; Gaut *et al.*, 2000). Similarly, the *a1* and *sh2* loci are separated by 140 kb in maize, but by only about 20 kb in sorghum and rice; most of the difference appears due to retrotransposon nests in maize (Chen *et al.*, 1998). Comparison of the LTR sequences indicated that the retrotransposons that expanded the *adh1* region in maize integrated between two and six million years ago, following the divergence of maize and sorghum 15 to 20 million years ago (SanMiguel *et al.*, 1998). Using similar estimation methods, Wei *et al.* (2002) estimated that the *Mla* locus of barley was invaded by five *BARE*-1 elements over the last 2 million years.

A major issue raised by evidence of massive genome expansion through integration of retrotransposons is the unidirectionality of the process: do genomes expand without end (Bennetzen and Kellogg, 1997a,b) Processes such as unequal crossing-over and deletion may remove retrotransposons, but do not appear sufficient alone to reverse the process of expansion. A third mechanism, LTR – LTR recombination, may play a major role. This removes one LTR and the internal domain of a retrotransposon from the genome and leaves behind a solo LTR. The genomes of *Hordeum* species, but apparently not that of maize, has 7 to 42 –fold more LTRs than full-length retrotransposons, corresponding in one barley cultivar to a total of 6×10^4 solo LTRs (Vicient *et al.*, 1999b). A striking contrast between the nests of retrotransposons observed in maize (SanMiguel *et al.*, 1996) and barley (Shirasu *et al.*, 2000) is that those in barley are comprised of solo LTRs but the nests in maize consist of full-length elements. In the compact yeast genome, likewise, 85% of the retroelements are present as solo LTRs (Kim *et al.*, 1998). The capacity of LTRs to recombine not only within a single retrotransposon, but also between elements along a chromosome arm, may well be a selective force for the gene island, repeat sea organization of cereal genomes. When LTRs from different elements recombine, the intervening DNA segment is lost. Selection would be unlikely to favor loss of genes by such a process, leading to segregation of retrotransposons and genes into different genomic compartments.

The dynamics of retrotransposon gain and loss in various regions of the genome has produced, much as for natural archipelagos, an array of both small and large gene islands separated by expanses of retrotransposons of varying sizes. These variations are reflected in differences in gene density observed in various contigs. For barley, Panstruga *et al.* (1998) and Shirasu and colleagues (2000) found a density of one gene per 20 kb overall, with

gene islands reaching one gene per 6 kb, Wei and coworkers (2002) observed average and island densities respectively of one gene per 8.1 kb and 4.6 kb, whereas Rostocks *et al.* (2002) reported an average of one gene per 21 kb, with variation from one gene per 12 kb to one gene per 103 kb. In maize, average densities of one gene per 50 kb, derived from islands of one per 9.8 kb surrounded by retrotransposon expanses of 14 to 70 kb, have been reported (Tikhonov *et al.*, 1999), though the *bz* locus and zein cluster reach densities as high as one gene per 3.2 kb and 5.2 kb respectively (Llaca and Messing, 1998; Fu *et al.*, 2001). Small gene islands of densities higher than one gene per 6 kb have been reported for barley, wheat, and rice as well (Feuillet and Keller, 1999), approaching the density of one gene per 4.5 kb found in the retrotransposon-poor *Arabidopsis thaliana.* The variability between different maize cultivars in the number of CentA elements in centromeres and retrotransposons in chromosomal knobs indicates that these regions, too, are subject to retrotransposon invasion and dynamic change over time (Ananiev *et al.*, 1998a,b). Currently, we do not know if these widely varying densities represent a snapshot of a process of ongoing division of gene islands by newly forming retrotransposon nests, or if accumulation of retrotransposons is selected against at some loci in cereal genomes.

Retrotransposons depend, for their evolutionary survival, on being active enough transcriptionally that sufficiently many functional copies will be inherited to insure against loss and mutational inactivation. Insertion in silent heterochromatin does not favor this goal, but reducing host fitness through integrating into genes seems counterproductive as well. Integrating into nests near functional genes, however, may offer retrotransposons a solution. The specific means, if they exist, in cereal or other plant genomes to target many retrotransposons into nests of other retrotransposons remain to be defined. However, in the compact yeast genome, mechanisms have evolved to target integration into narrowly defined regions (Zhou *et al.*, 1996; Boeke and Divine, 1998). The *Ty3* retrotransposon, for example, is directed to its insertion site near tRNA genes by the affinity of its integrase to transcription factors of RNA polymerase III, which transcribes tRNA (Kirchner *et al.*, 1995). The general model of nests of retrotransposons built up over evolutionary time either through insertional preference or post-integration selection appears, however, to concern mainly the families of prevalent elements, present in thousands of copies. The *Tnt1* element is present in several hundred copies in the tobacco genome, and was originally isolated because of its ability to be active in tissue culture and mutate genes (Grandbastien *et al.*, 1989). The *Tos17* element is present in only one to four copies in the rice genome, but it is sufficiently activated by tissue culture where it is highly mutagenic to make it an excellent tool for gene tagging

(Hirochika *et al.*, 1996; Yamazaki *et al.*, 2001). Recent data for 20,000 integration sites show that *Tos17* inserts three time more frequently into genic regions than into intergenic regions, preferring as its insertion site the palindrome ANGTT – AACNT surrounding the target duplication, and that 76% of these motifs are found in genic regions in the rice genome (Miyao *et al.*, 2003). The *Tos17* element is also transiently activated by the crossing of wild rice (*Zamia latifolia*) with cultivated rice, leading to a stable increase in copy number in the succeeding introgression lines (Liu and Wendel, 2000).

Little or no information exists on the target site specificities of the non-LTR retrotransposons, the SINEs and LINES, in cereals. Intriguingly, SINE elements in Brassica show a preference for integrating into matrix attachment regions (MARs) although not strict sequence specificity (Tikhonov *et al.*, 2001). The authors propose two possible explanations: MARs are cleavage targets for the LINE endonucleases likely involved in SINE integration; insertion into MARs may lead to changes in chromatin organization, affecting gene expression and generating useful variation.

A general view has developed that abundant elements, which may be, as is *BARE*-1, active in many tissues of the plant (Suoniemi *e al.*, 1996a; Jääskeläinen *et al.*, 1999; Vicient *et al.*, 2000b), have been able to achieve high copy number because they generally do not disrupt genes, whereas other elements such as *Tos17* are under tight regulation and consequently rare because of their mutagenic potential. In this regard, the MITE elements, which are not retrotransposons but rather are derivatives of nonautonomous Type II transposons (Jiang *et al.*, 2003), offer a contrasting model. The MITE elements are also present in high copy numbers in the genome, but show a strong propensity for inserting in or near genes (Bureau et al., 1996; Casa *et al.*, 2000). One of the obvious differences between MITEs and retrotransposons is their size, 100 to 300 bp for MITEs and up to 10 kb for retrotransposons. It is tempting to speculate that MITEs are tolerated within genes because of their low disruptive potential (Walbot and Petrov, 2001).

Both MITEs and prevalent retrotransposons have been shown, however, to affect gene activity (Wessler *et al.*, 1995). The simplest cases are loss of function mutations caused by integration into individual genes (*e.g.*, Weil and Wessler, 1990; Varagona *et al.*, 1992). Larger-scale changes in gene expression may be possible through SINE insertions at MARs (Tikhonov *et al.*, 2001). Transposable elements may also insert into or near promoters, contributing novel regulatory functions (Bureau *et al.*, 1994; White *et al.*, 1994; Kloeckener-Gruissem and Freeling, 1995). A spectacular example of global effects on gene expression is the transcriptional activation of the *Wis*-2 retrotransposons in wheat following a wide cross, and subsequent silencing

of hundreds of nearby genes by chimeric transcripts originating from the LTRs (Kashkush *et al.*, 2003). If phenotypic variability provides a basis for selection, then retrotransposons that integrate into genes and affect their expression, rather than those that insertions solely into silent seas of repetitive DNA, will at times have a selective advantage.

2.6. Retrotransposon-Based Molecular Markers: A Practical Use of Genome Organization

Retrotransposons, as major agents of genome change and also as major, dispersed components of the genome, are appealing candidates for the development of molecular marker systems designed to track such changes. The transposition of retrotransposons is not linked, as it is with DNA transposons, with removal of the mother element from its locus; although LTR-LTR recombination may delete an element, the remaining LTR leaves many such losses transparent for marker systems based on LTRs. The greater unidirectionality of retrotransposon integration compared to that of point mutations, or to microsatellite expansion and contraction, confers great advantages in reconstructing pedigrees and phylogenies. Furthermore, the ancestral state of a retrotransposon insertion is obvious – it is the empty site, whereas for most genetic polymorphisms on which markers are built it cannot be inferred. In this way, SINE elements have been used to trace human roots to Africa (Batzer *et al.*, 1994), to establish the relationship of whales to even-toed ungulates (Shimamura *et al.*, 1997), and to infer the evolutionary relationships between wild rice species (Cheng *et al.*, 2002).

Most retrotransposon-based methods employ PCR primed on conserved motifs in the element and on some other widespread and conserved motif in the surrounding DNA. Waugh and colleagues (1997) exploited the dispersion and prevalence of *BARE*-1 in barley through modification of the AFLP (amplified fragment length polymorphism) technique. In their approach, coined S-SAP (sequence-specific amplified polymorphism), a primer anchored in the LTR replaces one of the adapter primers of AFLP. As discussed above, although retrotransposons are dispersed, they are also clustered in the genome. It is this phenomenon that makes possible the IRAP (inter-retrotransposon amplified polymorphism) method, in which amplifications are carried out between primers for two retroelements in the genome (Kalendar *et. al.*, 1999). If *BARE*-1 elements were fully dispersed in the barley genome, amplification templates would be a minimum of 50 kb long. Likewise, the proximity of microsatellites to retrotransposons in cereal genomes (see below) makes the REMAP method (Kalendar *et al.*, 1999)

function. In addition to S-SAP, IRAP, and REMAP, a fourth based method based on the polymorphic integration pattern of retrotransposons, RBIP (retrotransposon-based insertional polymorphism), has been developed (Flavell *et al.*, 1998). The RBIP method uses primers flanking retrotransposon insertions and scores the presence or absence of insertions at individual sites. Over the last five years, retrotransposon-based marker systems have proven their utility in phylogenetic, genetic diversity, breeding, and mapping projects not only in cereals including barley, wheat, rye, and oat, but also in other crop plants and tree species (*e.g.*, Ellis *et al.*, 1998; Gribbon *et al.*, 1999; Manninen *et al.*, 2000; Yu and Wise, 2000; Boyko *et al.*, 2000; Porceddu *et al.*, 2002; Leigh *et al.*, 2003; also see the chapter no. 3 of this volume).

3. ORGANIZATION OF MICROSATELLITES

Microsatellites are tandem repeats of DNA sequences, each repeat only a few base pairs (1-6 bp) long (Tautz and Renz, 1984). These are more popularly described as simple sequence repeats (SSRs) in plant systems (Morgante and Oliveri, 1993) and as short tandem repeats (STRs) in animal systems (Edwards *et al.,* 1991). These motifs consist of a single base pair or a small number of bases (usually ranging from 1 to 6) which are repeated several times and known as di-, tri-, tetra-nucleotide repeats, etc., accordingly. Microsatellites are abundant and occur frequently and randomly in all eukaryotic DNAs examined so far (see Gupta *et al.*, 1996; Gupta and Varshney, 2000). In the past, on an average, microsatellites in plant genomes were shown to be 10 fold less frequent than in the human genome (Powell *et al.*, 1996), although more recently, the frequencies were shown to be comparable in plants and animal systems. For instance, according to an earlier estimate on the basis of database search at that time, on an average, in monocots one SSR was observed every 64 kb, while in dicots one SSR occurred every 21kb (Wang *et al.*, 1984). However, according to recent reports based on searching of large genomic and EST (expressed sequence tags) sequences in different species, the frequencies of SSRs in plant genomes were observed to be much higher (1 SSR every 6-7 kb) than those reported earlier (Cardle *et al.*, 2000; Varshney *et al.*, 2002; Morgante *et al.*, 2002), and were comparable to those described for mammals (Beckmann and Weber, 1992). Also, in the rice genome sequences released as rough draft by Syngenta, one SSR (only di-, tri- and tetra-nucleotide repeats) was found to occur every 8 kb (Goff *et al.*, 2002).

Variation was also observed in the frequencies of individual microsatellite motifs among different organisms (Lagercrantz *et al.*, 1993; Morgante and Oliveri, 1993; Wang *et al.*, 1994; Gupta *et al.*, 1996). For instance, $(CA)_n$ motif is one of the most frequently occurring microsatellite in humans and several other mammals, but is comparatively less frequent in plants (Lagercrantz *et al.*, 1993; Morgante *et al.*, 2002). In plants, however, $(AT)_n$ microsatellite is the most abundant and the $(GA)_n$ is relatively more abundant than $(CA)_n$ repeats (Wang *et al.* 1994; Gupta *et al.*, 1996; Morgante *et al.*, 2002).

3.1. Densities/ Frequencies and Organization of Microsatellites in the Whole Genome

During the last decade and the early years of the present decade, microsatellite markers were developed in a large number of plant systems including major cereal species such as barley (Ramsay *et al.*, 2000; Thiel *et al.*, 2003), maize (Chin, 1996; Yu *et al.*, 2001), oats (Li *et al.*, 2000), rice (Akagi *et al.*, 1996; Temnykh *et al.*, 2000, 2001; McCouch *et al.*, 2002; Gao *et al.*, 2003), rye (Saal and Wricke, 1999), sorghum (Bhattramakki *et al.*, 2000) and wheat (Roder *et al.*, 1998b; Varshney *et al.*, 2000; Gupta *et al.*, 2002). In the majority of these studies involving several plant species, the two most common SSRs, whose densities in the genome were determined while screening genomic libraries for SSRs, included GA/CT and GT/CA (Table 1). The densities of $(AT)_n$ or $(GC)_n$ motifs could not be worked out, because of the difficulty in hybridization, due to self annealing. The densities of GA/CT and GT/CA in different plant species, determined as above, ranged from one SSR every 212 kb to 704 kb. These results differ from those for humans, with an estimated average density of one SSR every 6 kb (Beckmann and Weber, 1992). Estimates of the total number of SSRs at the genome level have also been made in several crops. For instance, the frequencies per haploid genome were found to be 3.6×10^4 $(GA)_n$ and 2.3×10^4 $(GT)_n$ in bread wheat (Roder *et al.*, 1995), and were estimated to be 1.36×10^3 $(GA)_n$ and 1.23×10^3 $(GT)_n$ in rice (Panaud *et al.*, 1995).

The organization of microsatellites in plant genomes has been studied in the past using both hybridization-based and PCR-based approaches. In hybridization-based approaches, synthetic oligonucleotides are used as probes either for hybridizing the gels or filters containing genomic DNA digested with a restriction enzyme, or for *in situ* hybridization of chromosomes. In PCR approaches, on the other hand, the primers flanking the microsatellites are designed and used in mapping/tagging experiments.

Table 1. Density of microsatellites in different portions of cereal genomes

Crop and source of SSRs	Density (kb of DNA per SSR)	Reference
Barley		
Genomic DNA	7.4	Cardle *et al.* (2000)
ESTs	7.5	Varshney *et al.* (2002)
	3.4	Kantety *et al.* (2002)
Maize		
Genomic DNA	4.5/5.71	Morgante *et al.* (2002)*
ESTs	8.1	Cardle *et al.* (2000)
	1.63/2.12	Morgante *et al.* (2002)*
	1.5	Kantety *et al.* (2002)
	7.5	Varshney *et al.* (2002)
	28.32	Gao *et al.* (2003)
Rice		
Genomic DNA	225-240	Wu and Tanksley (1993)§
	330-365	Panaud *et al.* (1995)¶
	7.4	Cardle *et al.* (2000)
	16/1.9	Temnykh *et al.* (2001)
	2.64/3.52	Morgante *et al.* (2002)*
BAC end sequences	40/3.7	Temnykh *et al.* (2001)
ESTs	3.4	Cardle *et al.* (2000)
	19	Temnykh *et al.* (2001)
	0.86/1.06	Morgante *et al.* (2002)*
	3.9	Varshney *et al.* (2002)
	4.7	Kantety *et al.* (2002)
	11.81	Gao *et al.* (2003)
Rye		
ESTs	5.5	Varshney *et al.* (2002)
Sorghum		
ESTs	5.5	Varshney *et al.* (2002)
	3.6	Kantety *et al.* (2002)

Table 1. Continued

Wheat		
Genomic DNA	440-704	Roder *et al.* (1995)[§]
	212-292	Ma *et al.* (1996)[§]
	3.35/5.16	Morgante *et al.* (2002)[*]
ESTs	1.33/1.67	Morgante *et al.* (2002)[*]
	6.2	Varshney *et al.* (2002)
	3.2	Kantety *et al.* (2002)
	17.2	Gao *et al.* (2003)
	9.2	Gupta *et al* (2003)

[*]analysed frequency of imperfect and perfect SSRs separately. In each case frequency of imperfect SSRs is given first and followed by that of perfect SSRs; [§]studied the frequency of two DNR SSRs (GA/CT, GT/CA); [¶]classified SSRs in two categories- Class I SSRs >20 bp and ClassII SSRs >12 bp <20 bp. In each case the frequency of class I SSR is followed by that of class II.

Based on studies using the above approaches, conclusions have been drawn about the organization of microsatellites within a genome, although one should recognize the limitation of the hybridization approaches, which generally detect only high-density fragments representing the repetitive DNA.

3.1.1. *In-Gel* Hybridization (Oligonucleotide Fingerprinting)

The synthetic oligonucleotide probes complementary to SSR motifs have been successfully utilized for in-gel hybridization (and sometimes for Southern hybridization) with genomic DNA that was digested with individual restriction enzymes and electrophoresed on agarose gels (Ali *et al.*, 1986). The fragments that hybridize with synthetic oligonucleotides through in-gel hybridization are generally many, and the size of the hybridizing fragments range from a few hundred base pairs to more than 20kb, thus making the technique suitable for DNA fingerprinting (Beyermann *et al.*, 1992; Schmidt and Heslop-Harrison, 1996; Arens *et al.*,1995; Weising *et al.*, 1995, 1998). In crops like wheat, however, multilocus fingerprints due to SSR probes that are characteristic of this technique were not obtained during in-gel hybridization. Instead, a prominent solitary high molecular weight fragment (>23 kb), sometimes associated with a few low molecular weight bands, was obtained with a number of SSR probes (Varshney *et al.*, 1998). Similar type of high high molecular weight fragments (>23 kb) were observed in 14 individual species of *Triticum-Aegilops* group after in-gel hybridization (Sharma *et al.*, 2002). The presence of high molecular weight fragments measuring up to 30 kb and containing

$(GATA)_n$ and $(GTG)_n$ stretches were also earlier reported in barley (Beyermann *et al.,* 1992). In sugar-beet, also, fragments of >21.2 kb were obtained after hybridization with $(CA)_8$ probes; suggesting that tandem repeats harboring $(CA)_n$ occur in genomic regions, and that the organization of $(CA)_n$ differs from all the other microsatellites used in this study (Schmidt and Heslop-Harrison, 1996). It is also known that stretches carrying $(CA)_n$ form the subrepeats of a centromeric satellite (Schmidt and Metzlaff, 1991). In our opinion, the presence of high molecular weight bands (>20 kb) hybridizing with SSRs indicates the occurrence of microsatellites either within or in close association with the long tandem repeat units, *e.g.* retrotransposons, etc. Since the technique of in-gel hybridization is suitable for detection of only those sequences, which are relatively long and repetitive, the fragments that hybridized with synthetic oligonucleotides during in-gel hybridization may actually represent only those DNA fragments associated with SSRs, which are generally repetitive rather than representing both the repetitive and the unique sequences (Varshney *et al.*, 1998).

3.1.2. Association of Microsatellites with Retrotransposons

SSRs have been shown to be associated with 'short interspersed elements' (SINEs) in rice (Motohashi *et al.*, 1997). In barley also, association of microsatellites was observed with retrotransposons and other dispersed repetitive elements like *BARE-1, WIS2-1A, R-173, Pgr-1/PREM-1* (Ramsay *et al.*, 1999). In another study in rice, about 45% of $(AT)_n$ SSRs (harbored in BAC-ends) showed significant homology to 'Micropan' sequences, a new family of 'miniature inverted-repeat transposable elements' (MITEs) (Temnykh *et al.*, 2001). On the basis of the relative position of SSR and retrotransposons in barley, Ramsay *et al.* (1999) has drawn inferences about the origin of this association. They postulated that while some SSRs (proto-SSR; A- rich sequence) might have acted as 'landing pads' for insertion of transposable elements, there may be other SSRs that must have evolved as components of active transposable elements that are spread throughout the genome. It is also possible that, in some cases, expansion of SSRs in retroelements might have led to multiple SSR loci, which may also account for the difficulty in locus specific amplification, while using locus specific STMS primers. This may be true of the genomes like that of wheat, which contains a higher proportion of repetitive DNA (Varshney, 2001). The association between SSRs and mobile elements, as above, has also facilitated the development of the novel marker system REMAP (Kalender *et al.*, 1999; see section 2.6 earlier).

3.1.3. *In Situ* Hybridization with SSR Probes

In situ hybridization (ISH) was initially used to assess the chromosomal localization of various SSRs in humans and animals (Pardue *et al.*, 1987; Lohe *et al.*, 1993). This gave some useful information regarding physical organization of microsatellites in plant chromosomes, including those from cereals. For instance, a $(GAA)_7$ probe, when hybridized to barley chromosomes, gave a pattern conforming to the distribution of heterochromatin and the C-banding pattern (Pederson and Linde-Laursen, 1994). This distribution of the so-called (GAA) satellite in heterochromatic regions was later confirmed in some other cereal species also (Pederson *et al.*, 1996). In bread wheat, rye, and hexaploid triticale, ten different SSRs gave dispersed hybridization signals of varying strength on all chromosomes when used for ISH (Cuadrado and Schwaracher 1998). However, in wheat, microsatellite motifs $(AG)_{12}$, $(CAT)_5$, $(AAG)_5$, $(GCC)_5$, and in particular, $(GACA)_4$ hybridized strongly to pericentromeric and multiple intercalary sites on the B genome chromosomes and on chromosome 4A, resembling the N-banding pattern. In contrast to this, in rye, $(GACA)_4$ gave strong hybridization signals at many intercalary sites in all the chromosomes, which largely differed from the known banding pattern. Earlier in sugar beet also, when CA, GA, TA, CAC, GATA, GACA and GGAT repeats were used for ISH, it was noticed that each microsatellite had a characteristic genomic distribution and motif-dependent dispersion with site-specific enrichment or depletion of some motifs at centromeric or intercalary positions (Schmidt and Heslop-Harrison, 1996). From several ISH studies, as above, it has been inferred that the tandemly repeated sequences are located around all centromeres in blocks, more than 80 kb long, and that their homology in the microsatellite domain perhaps is responsible for some of the hybridization signals. Thus, ISH studies support the hypothesis of association of at least some SSRs with highly repetitive DNA that may sometimes involve retrotransposons.

3.1.4. Genome Mapping and Physical Mapping using STMS Primer Pairs

The sequences flanking specific microsatellite loci in the genome are believed to be conserved and, therefore, have been used for designing primers to amplify individual microsatellite loci: the technique was described as sequence tagged microsatellite site (STMS) analysis (Beckmann and Soller, 1990). STMS primers have come to be used extensively for genome mapping in a number of plant species including wheat, barley, rice, maize, etc. (for references see Gupta and Varshney, 2000). In almost all plant

species, the genome mapping suggested that microsatellites are dispersed throughout the genome and not clustered (Gupta and Varshney, 2000) except some reports in barley (Ramsay *et al.*, 2000; Li *et al.*, 2003), where centromeric clustering of genomic SSRs was observed. No clustering in barley genome was, however, observed, when EST-derived microsatellite loci were mapped later (Thiel *et al.*, 2003, Varshney *et al.*, unpublished). In rice, microsatellite markers with different SSR motifs, regardless of whether they belonged to genomic DNA or cDNA, were also found to be rather uniformly distributed along all the rice chromosomes (Panaud *et al.*, 1996; Temynkh *et al.*, 2000; McCouch *et al.*, 2002). Similarly, in bread wheat, physical mapping of microsatellite markers on chromosomes of homoeologous group 2 showed an absence of microsatellite clustering (Roder *et al.*, 1998a).

3.1.5. Frequencies of Microsatellites in Whole Rice Genome

The distribution of SSRs in the rice genome has also been studied on the basis of the two whole genome draft sequences released, respectively, by Syngenta and by the Beijing Genome Institute (BGI). In the draft sequence released by Syngenta (Goff *et al.*, 2002), for instance, 48,351 SSRs (including di-, tri- and tetra-nucleotide repeats) were available, giving a density of 8 kb per SSR in the whole genome; SSRs represented by di-, tri-, and tetra-nucleotide repeats accounted respectively for 24%, 59% and 17% of the total SSRs. The most frequent dinucleotide repeats (DNRs) were AG/CT repeats, which accounted for 58% of all DNRs, and the most frequent trinucleotide repeats (TNRs) were CGG/CCG repeats, which accounted for 44% of all TNRs. In the predicted genes, among the 7000 SSRs that were available, SSRs were mainly TNRs (92%). This abundance of TNRs may be attributed to a lack of selection against length variation in these SSRs, since it will not cause any frameshift mutations. SSRs were also studied in the rough draft of rice genome sequences released by BGI (Yu *et al.*, 2002), and accounted for 1.7% of the genome, as against 3% of the human genome represented by SSRs. Interestingly, majority of rice SSRs were mononucleotides, primarily $(A)_n$ or $(T)_n$.

3.2. Density and Distribution of Microsatellites based on *in silico* Mining

During the last five years, in the genomics era, large-scale genome/EST sequencing projects were initiated in several plant species including cereals.

The data generated from these projects was utilized for studying the frequency, distribution and organization of microsatellites in the expressed portion of the genome, and in some cases also in the whole genome (Table 1). For development of EST-SSRs, ESTs have been scanned in different plant species, including cereals such as rice (Temnykh *et al.*, 2000, 2001), barley (Kota *et al.,* 2001; Thiel *et al.*, 2003), wheat (Eujayl *et al.*, 2002; Gao *et al.*, 2003; Gupta *et al.*, 2003), and rye (Hackauf and Wehling, 2002). These efforts also allowed estimation of the density of SSRs in expressed regions of the genomes. For instance, on the basis of a number of contiguous genomic sequences, the density of SSRs was found to be 7.4 kb per SSR in barley and 7.4 kb per SSR in rice. Similarly, on the basis of EST sequences, the density of SSRs was 3.4 kb per SSR in rice and 8.1 kb per SSR in maize (Cardle *et al.*, 2000). Another study was conducted, which involved survey of EST sequences amounting to 75.2 Mb in barley, 54.7 Mb in maize, 43.9 Mb in rice, 3.7 Mb in rye, 41.6 Mb in sorghum and 37.5 Mb in wheat; the overall average density of SSR in these species was found to be 6.0 kb per SSR (Varshney *et al.*, 2002). However in another study, the frequency of SSR was one every 11.81 kb in rice, 17.42 kb in wheat and 28.32 kb in maize (Gao *et al.*, 2003). Difference in the frequency of SSRs in the ESTs of a particular species in different studies may be attributed to criteria of SSR search and data quantity used to identify SSRs in the database mining approaches.

In the comprehensive study of Varshney et al. (2002), almost in every cereal species, the TNRs were the most frequent (54% to 78%) followed by the DNRs (17.1% to 40.4%). As mentioned earlier, the abundance of trimeric SSRs was attributed to the absence of frameshift mutations due to length variation in these SSRs (Metzgar *et al.*, 2000). Among the trinucleotide repeats also, codon repeats corresponding to small hydrophilic amino acids are perhaps easily tolerated, but strong selection pressures probably eliminate codon repeats encoding hydrophobic and basic amino acids. Such an inference was drawn from an analysis of coding DNA sequences in the whole genomes of fruitfly, the nematode *C. elegans* and the budding yeast (Katti *et al.*, 2001). Furthermore, in various cereal genomes, among the DNRs, the motif AG is the most frequent (38% to 59%) followed by the motif AC (20% to 34%) in all the species except rye, where these frequencies are 50% for AC and 37.9% for AG (Varshney *et al.*, 2002). The most infrequent motif is CG in all species (1.7% to 9.0%) except in barley, where AT is the least frequent (8.4%). Among the TNRs, the motif CCG is the most frequent, ranging from 32% in wheat, to 49% in sorghum followed by AGC (13% to 30%) in barley, maize, rice and sorghum, and AAC in wheat (27%) and rye (16%). The third most frequent motif is AGG in barley, rice, rye, sorghum, AGC in wheat, and AAC in maize.

In another study of SSRs in rice genome, a total of 57.8 Mb DNA sequences were used, which included the following: (i) 12,532 ESTs (6.3 Mb), (ii) a large set (74,127) of short BAC-end sequences (500 bp on average), and (iii) 27 fully sequenced large-insert (BAC/PAC) clones (150 kb on average) (Temynkh *et al.* 2000, 2001). In this study, SSRs were classified in two categories: class I SSRs, each with ≥20 bp, and class II SSRs, each with ≥12 bp to <20 bp. The density of class I SSRs was one SSR every 40 kb in BAC-end sequences, one SSR every 16 kb in fully sequenced BAC/PAC clones and one SSR every 19 kb in ESTs. Class II SSRs, similarly, occurred every 3.7 kb in BAC-ends and every 1.9 kb in BAC/PAC clones (Table 1). The proportion of GC-rich TNRs among the total SSRs also differed in each of the three classes of DNA sequences used, from a low of 10.5% of all SSRs in BAC-ends to a high of 59% of all SSRs in ESTs with intermediate frequencies (~27%) in fully sequenced BAC and PAC clones. The densities of $(AT)_n$ DNRs were in the reverse order to that of GC-rich TNR e.g. 38.2% in BAC ends, 27% in BAC/ PAC clones, and only 2.9% in ESTs. The pattern of the frequencies of tetra-nucleotide SSRs and (CA)n DNRs were similar to that of (AT)n.

The above pattern of variation demonstrated that different SSR motifs are not randomly distributed in the rice genome. Since the frequencies of GC-rich TNRs and (AT)n DNRs were identical in BAC/PAC clones and varied greatly and inversely in BAC-ends and ESTs, we may conclude that the BAC-ends and the ESTs represented different genomic domains, neither of which would have the SSR composition representative of the whole genome. However, the frequencies of different SSRs in the fully sequenced BACs/PACs perhaps represent more faithfully the frequencies of microsatellites in the whole genome. As a result, it can be concluded that regions of relatively high gene-density (e.g. BAC/PAC clones) have prevalence of GC-rich TNRs (known to be associated with genes) and the regions of lower gene density (e.g. BAC ends) has prevalence of (AT)n, (AC)n and tetra-nucleotide SSRs (which are abundant in non-coding, intergenic regions).

3.3. Comparative Distribution of Microsatellites in Transcribed and Non-transcribed Portions of the Genome

Recently, the abundance and relative distribution of microsatellites among transcribed and non-transcribed regions have been more critically assessed in some plant species including rice, maize and wheat, after examining a large data set of genomic sequences and ESTs (Morgante *et al.*, 2002). In general,

the frequency of microsatellites was significantly higher in ESTs than in genomic DNA across all species (Table 1). In the maize genome as well, the frequency of microsatellites in the non-repetitive fraction was actually found to be significantly higher (more than double) than that in the repetitive fraction. Furthermore, the average density of microsatellites in rice BAC clones that were representative of gene rich regions was almost double that of the reference rice genomic DNA, which again supports the hypothesis of higher frequencies of microsatellites in the non-repetitive regions. A highly significant and positive linear relationship was also observed between frequencies of microsatellites (including both, the perfect and the imperfect SSRs) and the proportion of single copy DNA in a genome. Thus microsatellite frequency is a function of not only the overall genome size, but also of the relative proportion of single-copy DNA.

The above results are in sharp contrast to the earlier observations, which assumed an occurrence of SSRs mainly in the repetitive DNA, and therefore, were used to suggest preferential origin of microsatellites from repetitive DNA in both animals (Arcot *et al.*, 1995; Nadir *et al.*, 1996) and plants (Ramsay *et al.*, 1999). However, these earlier results can now be explained on the basis of the limitation of the techniques, which perhaps failed to score SSRs in the unique sequences. Frequencies of individual SSRs also varied both within and between genomic and transcribed sequences. For instance, in ESTs there was a higher frequency of AG/CT repeats and lower frequency of AT repeats. Similarly, TNRs were significantly more abundant in ESTs and the CCG/CGG repeat motifs alone accounted for half of the TNRs in rice ESTs. In contrast to this, they were rare in dicots (e.g. Arabidopsis and soybean) and moderately abundant in monocots other than rice (e.g. maize and wheat). This difference was attributed to higher G+C content and consequent usage bias in monocot ESTs. Overall, on an average, microsatellites were also shorter in the transcribed regions.

On the basis of the *in silico* analysis of a large data set in various species and the rough drafts of rice genome that were released, the SSR frequency in plant, or at least in cereal, genomes is in the range of one SSR every 2- 10 kb, which is comparable with those for humans and mammals (Beckmann and Weber, 1992). Further, it can be concluded clearly that TNRs comprise the highest proportion of the total SSRs followed by the DNRs. The most frequent motifs are CCG, AGC, AAC among TNRs, and AG and AC among the DNRs. However, there are discrepancies in the frequencies of SSRs in a particular class within a given species (Table 1). These discrepancies may be explained by the varying search criteria for SSRs or by the different bias to 5' or 3' regions in the ESTs or by varying dataset, used by different groups. Among different species, the overall frequency of microsatellites is inversely

related to genome size and to the proportion of repetitive DNA but remained constant in the transcribed portion of the genome. This indicates that the distribution of microsatellites is a function of the dynamics and history of genome evolution and of selective constraints, because these microsatellites resided in different regions of the genome of a species pre-dating the recent genome expansions in the plants (Morgante *et al.*, 2002).

4. SUMMARY AND OUTLOOK

The wide variations in genome size, independent of biological complexity, phylogenetic relationship, or ploidy level, that were described as the *C*-value paradox in the early 1970s are now known to be due to the variations in abundance of the retrotransposons, DNA transposons, simple sequence repeats, and other repeats that constitute the greater portion of most plant genomes. The genome, rather than being a static association of genes recombining with a frequency based on their physical distance and accumulating point mutations, is a complex and dynamic landscape in which these various elements create both local and large-scale structures that change over time. The presence in cereals of simple sequence repeats in and near retrotransposons and genes, and the insertion of retrotransposon into the blocks of repetitive elements present in cereal centromeres, represent just a part of the landscape that genome sequencing is now revealing.

The retrotransposons, far from being the "junk DNA" they were initially in ignorance referred to as, function as self-replicating genomic viruses. They are able to effect vast changes in genome size over evolutionary time as well as alter gene expression patterns and gene products in single generations. Ultimately, the balance between gain and loss of retrotransposons, within the context of their integrational preferences and selective forces, shapes the genome. LTR-LTR recombination, gene conversion mechanisms, and the effects of repeated small deletions can offset gain through integration. At individual loci, reconstruction by means of the direct repeats generated by integrase reveals cycles of gain by integration and loss through LTR-LTR recombination (Shirasu *et al.*, 2000). Hot, dry conditions appear to favor gain but decrease loss of retrotransposons (Kalendar *et al.*, 2000). Loss of retrotransposons may in itself provide a purifying selection for functionality both of coding domains (Suoniemi *et al.*, 1998; Navarro-Quezada and Schoen, 2002) and of the promoter.

For the retrotransposons, successive cycles of nested integration, LTR-LTR recombination, and accumulation of point mutations and small deletions ultimately obliterates the historical record of events in the genome, limiting

our historical reach. The bidirectionality and rapidity with which microsatellites can change in size likewise obscures their origins. Although a small genome with few retrotransposons is generally seen as the ancestral state, proving this is difficult. Nevertheless, over the time scale represented by agriculture, 10 to 20 thousand years, retrotransposon insertions and microsatellite allelic variations are highly useful as molecular markers in a wide variety of applications. Mapping, phylogeny reconstruction, genetic diversity studies, pedigree analysis, and targeted breeding is proving the practical value of the "junk" and has reunited it with genetics.

5. ACKNOWLEDGEMENTS

We thank Dr Robbie Waugh, Scottish Crop Research Instititute, Invergowrie, Scotland for reviewing the MS and helpful comments.

6. REFERENCES

Abrahamson S., Bender M.A., Conger A.D., Wolff S. Uniformity of radiation-induced mutation rates among different species. Nature 1973; 245: 460-462

Akagi H., Yokozeki Y., Inagaki A., Fujimura T. Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 1996, 93:1071-1077

Ali S., Müller C.R., Epplen J.T. DNA fingerprinting by oligonucleotide probes specific for simple repeats. Hum Genet 1986; 74:239-243

Ananiev E.V., Phillips R.L., Rines H.W. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 1998a; 95: 13073-13078

Ananiev E.V., Phillips R.L., Rines H.W. Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics 1998b; 149: 2025-2037

Arcot S.S., Wang Z.Y., Weber J.L., Deininger P.L., Batzer M.A. Alu repeats- a source for the genesis of primate microsatellites. Genomics 1995; 29: 136-144

Arens, P., Odinot P., van Heusden A.W., Lindhout P., Vosman B. GATA and GACA repeats are not evenly distributed throughout the tomato genome. Genome 1995; 38:84-90

Arnaud P., Goubely C., Pélissier T., Deragon J.-M. SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol Cell Biol 2000; 20: 3434-3441

Baltimore D. RNA-dependent DNA polymerase in virons of RNA tumor viruses. Nature 1970; 226: 1209-1211

Barakat A., Carels N., Bernardi G. The distribution of genes in the genomes of Gramineae. Proc Natl Acad Sci USA 1997; 94: 6857-6861

Batzer M.A., Stoneking M., Alegria-Hartman M., Bazan H., Kass D.H., Shaikh T.H., Novick G.E., Ioannou P.A., Scheer W.D., Herrera R.J., *et al.* African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 1994; 91: 12288-12292

Beckmann, J.S., Soller M. Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Teclmology 1990; 8: 930-932

Beckmann J.S. Weber J.L. Survey of human and rat microsatellites. Genomics 1992; 12: 627-631

Bennetzen J.L., Freeling M. The unified grass genome: synergy in synteny. Genome Res 1997; 7: 301-306

Bennetzen J.L., Kellogg E.A. Do plants have a one-way ticket to genome obesity? Plant Cell 1997a; 9: 1509-1514

Bennetzen J.L., Kellogg E.A. Reply. Plant Cell 1997b; 9: 1901-1902

Bennetzen J.L., Schrick K., Springer P.S., Brown W.E., Sanmiguel P. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 1994; 37: 565-576

Beguiristain T., Grandbastien M.A., Puigdomènech P., Casacuberta J.M. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol 2001; 127: 212-221

Belyayev A., Raskina O., Nevo E. Chromosomal distribution of reverse transcriptase-containing retroelements in two Triticeae species. Chrom Res 2001; 9: 129-136

Beyermann B., Nurnberg P., Weihe A., Mexiner M., Epplen J.T., Borner T. Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive sequences. Theor Appl Genet 1992; 83: 691-694

Bhattramakki D., Dong J.M., Chhabra A.K., Hart G.E. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 2000; 43: 988-1002

Boeke J.D. LINEs and Alus -- the polyA connection. Nat Genet 1997; 16: 6-7

Boeke J.D., Garfinkel D.J., Styles C.A., Fink G.R. *Ty* elements transpose through an RNA Intermediate. Cell 1985; 40: 491-500

Boeke J.D., Devine S.E. Yeast retrotransposons: finding a nice quiet neighborhood. Cell 1998; 93: 1087-1089

Boyko E., Kalendar R., Korzun V., Gill B., Schulman A.H. Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers. Plant Mol Biol 2002; 48: 767-790

Bureau T.E., Wessler S.R. *Stowaway* : A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 1994; 6: 907-916

Bureau T.E., Ronald P.C., Wessler S.R. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci USA 1996; 93: 8524-8529

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 2000; 156: 847-854

Casa A., Brouwer C., Nagel A., Wang L., Zhang Q.-., Kresovich S., Wessler S.R. The MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. Proc Natl Acad Sci USA 2000; 97: 10083-10089

Casacuberta E., Pardue M.L. Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilus group. Proc Natl Acad Sci USA 2003; 100: 3363-3368

Chen M., SanMiguel P., de Oliviera A.C., Woo S.-S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431-3435

Chen M., SanMiguel P.J., Bennetzen J.L. Seguence organization and conservation in *sh2/a1* -homologous regions of sorghum and rice. Genetics 1998; 148: 435-443

Cheng C., Tsuchimoto S., Ohtsubo H., Ohtsubo E. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. Genes Genet Syst 2002; 77: 323-334

Chin E.C.L. Maize simple repetitive DNA sequences: abundance and allele variation. Genome 1996; 39:866-873

Cuadrado A., Schwarzacher T. The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 1998; 107:587-594

Deragon J.-M., Capy P. Impact of transposable elements on the human genome. Annals of Medicine 2000; 32: 264-273

Danilevskaya O.N., Lowenhaupt K., Pardue M.L. Conserved families of the *Drosophila* HeT-A telomere-specific retrotransposon. Genetics 1998; 148: 233-242

Dong F.G., Miller J.T., Jackson S.A., Song J., Jiang J. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 1998; 95: 8135-8140

Echenique V., Stamova B., Wolters P., Lazo G., Carollo L., Dubcovsky J. Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the Triticeae EST databases. Theor Appl Genet 2002; 104: 840-844

Edwards A., Civitello H., Hammond H.A. and Caskey C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991; 49: 746-756

Eickbush T.H. Transposing without ends: the non-LTR retrotransposable elements. New Biol 1992; 4: 430-440

Ellis T.H.N., Poyser S.J., Knox M.R., Vershinin A.V., Ambrose M.J. *Ty1-copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 1998; 260: 9-19

Eujayl I., Sorrells M.E., Baum M., Wolters P., Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 2002; 104: 399-407

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.*, Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Feuillet C., Keller B. High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 1999; 96: 8265-8270

Flavell A.J., Ruby S.W., Toole J.J., Roberts B.E., Rubin G.M. Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. Proc Natl Acad Sci USA 1980; 77: 7107-7111

Flavell R.B., Bennett MD, Smith J.B., Smith D.B. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 1974; 12: 257-268

Flavell A.J., Knox M.R., Pearce S.R., Ellis T.H.N. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant Jour 1998; 16: 643-650

Fu H., Park W., Yan X., Zheng Z., Shen B., Dooner H.K. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. Proc Natl Acad Sci USA 2001; 98: 8903-8908

Fukui K.-N., Suzuki A., Lagudah E.S., Rahman S., Appels R., Yamamoto M., Mukai Y. Physical arrangement of retrotransposon-related repeats in the centromeric regions of wheat. Plant Cell Physiol 2001; 42: 189-196

Gao L.F., Tang J.F., Li H.W., Jia J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 2003; 12: 245-261

Gaut B.S., Le Thierry d'Ennequin M., Peek A.S., Sawkins M.C. Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 2000; 97: 7008-7015

Gilbert N., Arnaud P., Lenoir A., Warwick S.I., Picard G., Deragon J.-M. Plant S1 SINEs as a model to study retroposition. Genetica 1997; 100: 155-160

Goodman H.M., Ecker J.R., Dean C. The genome of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 1995; 92: 10831-10835

Goff S.A., Ricke, D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma, H., Hadley, D. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *japonica*). Science 2002; 296:92-100

Grandbastien M.-A., Spielmann A., Caboche M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 1989; 337: 376-380

Grandbastien M.-A. Activation of retrotransposons under stress conditions. Trends Plant Sci 1998; 3: 181-187

Gribbon B.M., Pearce S.R., Kalendar R., Schulman A.H., Paulin L., Jack P., Kumar A., Flavell A.J. Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. Mol Gen Genet 1999; 261: 883-891

Gupta P.K., Balyan H.S., Edwards K.J., Isaac P., Korzun V., Roder M., Gautier M.F., Joudrier P., Schlatter A.R., Dubcovsky J., *et al.* Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 2002; 105: 413-422

Gupta P.K., Balyan H.S., Sharma P.C, Ramesh B. Microsatellites in plants: a new class of molecular markers. Curr Sci 1996; 70:45-54.

Gupta P.K., Rustgi S., Sharma S., Singh R., Kumar N., Balyan H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genom 2003; 270: 315-323

Gupta P.K., Varshney R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 2000; 113:163-165.

Hackauf B., Wehling P. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 2002; 121:17-25

Hake S., Walbot V. The genome of *Zea mays*, its organization and homology to related grasses. Chromosoma 1980; 79: 251-270

Heslop-Harrison J.S. Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. Plant Cell 2000; 12: 617-635

Hirochika H., Fukuchi A., Kikuchi F. Retrotransposon families in rice. Mol Gen. Genet. 1992; 233: 209-216

Hirochika H. Activation of tobacco retrotransposons during tissue culture. EMBO J 1993; 12: 2521-2528

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H., Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 1996; 93: 7783-7788

Hudakova S., Michalek W., Presting G.G., ten Hoopen R., dos Santos K., Jasencakova Z., Schubert I. Sequence organization of barley centromeres. Nucl Acids Res 2001; 29: 5029-5035

Jacob H.J., Brown D.M, Bunker R.K., Daly M.J., Dzau V.J., Goodman A., Kolike G., Kren V., Kurtz T., Lernmark A. *et al.* A genetic linkage map of the laboratory rat, *Rattus norvegicus*. Nature Genet 1995; 9: 63-69

Jääskeläinen M., Mykkänen A.-H., Arna T., Vicient C., Suoniemi A., Kalendar R., Savilahti H., Schulman A.H. Retrotransposon *BARE*-1: Expression of encoded proteins and formation of virus-like particles in barley cells. Plant Jour 1999; 20: 413-422

Jiang N., Bao Z., Zhang X., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. An active DNA transposon family in rice. Nature 2003; 421: 163-167

Kalender R., Grob T., Regina M., Suoniemi A., Schulman A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 1999; 98:704-711

Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 2000; 97: 6603-6607

Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48: 501-510

Kashkush K., Feldman M., Levy A.A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 2003; 32: 102-106

Katsiotis A., Schmidt T., Heslop-Harrison J.S. Chromosomal and genomic organization of *Ty1-copia*-like retrotransposon sequences in the genus *Avena*. Genome 1996; 39: 410-417

Katti M.V., Ranjeker P.K., Gupta V. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 2001; 18: 1161-1167

Kazazian H.H., Jr., Goodier J.L. LINE drive: Retrotransposition and genome instability. Cell 2002; 110: 277-280

Kim J.M., Vanguri S., Boeke J.D., Gabriel A., Voytas D.F. Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 1998; 8: 464-478

Kingsman A.J., Kingsman S.M. *Ty*: a retroelement moving forward. Cell 1988; 53: 333-335

Kirchner J., Connolly C.M., Sandmeyer S.B. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of retroviruslike element. Science 1995; 267: 1488-1491

Kloeckner-Gruissem B., Freeling M. Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. Proc Natl Acad Sci USA 1995; 92: 1836-1840

Komatsu M., Shimamoto K., Kyozuka J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. Plant Cell 2003; 15: 1934-1944

Kota R., Varshney R.K., Thiel T., Dehmer K.J., Graner A. Generation and comparison of EST-derived SSR and SNP markers in barley *(Hordeum vulgare* L.). Hereditas 2001; 135:145-151

Kumar A., Bennetzen J. Plant retrotransposons. Annu Rev Genet 1999; 33:479-532.

Kurata N., Nomomura K.-I., Harushima Y. Rice genome organization: the centromere and genome interactions. Ann Bot 2002; 90: 427-435

Kumekawa N., Ohtsubo E., Ohtsubo H. Identification and phylogenetic analysis of *gypsy*-type retrotransposons in the plant kingdom. Genes Genet Syst 1999; 74: 299-307

Kumekawa N., Hosouchi T., Tsuruoka H., Kotani H. The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. DNA Res 2000; 7: 315-321

Kumekawa N., Ohmido N., Fukui K., Ohtsubo E., Ohtsubo H. A new *gypsy*-type retrotransposon, *RIRE7*: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics 2001a; 265: 480-488

Kumekawa N., Hosouchi T., Tsuruoka H., Kotani H. The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. DNA Res 2001b; 8: 285-290

Lagercrantz U., Ellegren H., Andersson L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 1993; 21: 1111-1115.

Lamb J.C., Birchler J.A. The role of DNA sequence in centromere formation. Genome Biology 2003; 4: 214

Langdon T., Seago C., Mende M., Leggett M., Thomas H., Forster J.W., Jones R.L., Jenkins G. Retrotransposon evolution in diverse plant genomes. Genetics 2000; 156: 313-325

Laten H., Majumdar A., Gaucher E.A. SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 1998; 95: 6897-6902

Le Q.H., Wright S., Yu Z., Bureau T. Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 2000; 97: 7376-7381

Leigh F., Kalendar R., Lea V., Lee D., Donini P., Schulman A.H. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Genet Genom 2003; 269: 464-474

Lenoir A., Lavie L., Prieto J.-L., Goubely C., Côté J.-C., Pélissier T., Deragon J.-M. The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. Mol Biol Evol 2001; 18: 2315-2322

Levis R., Dunsmuir P., Rubin G.M. Terminal repeats of the Drosophila transposable element *copia*: nucleotide sequence and genomic organization. Cell 1980; 21: 581-588

Li C., Rossnagel B.G., Scoles, G. J. The development of oats microsatellite markers and their use in identifying relationships among *Avena* species and oats cultivars. Theor Appl Genet 2000; 101:1259-1268

Li J.Z., Sjakste T.G., Röder M.S., Ganal MW. Development and genetic mapping of 127 new microsatellite markers in barley. Theor Appl Genet 2003; 107: 1021-1027

Linares C., Serna A., Fominaya A. Chromosomal organization of a sequence related to LTR-like elements of Ty1-*copia* retrotransposons in *Avena* species. Genome 1999; 42: 706-713

Liu B., Wendel J.F. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 2000; 43: 874-880

Llaca V., Messing J. Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. Plant Jour 1998; 15: 211-220

Lohe A.R., Hilliker A.J., Roberts P.A. Mapping simple repeated DNA-sequences in heterochromatin of *Drosophila melanogaster*. Genetics 1993; 134:1149-1174

Löwer R., Löwer J., Kurth R. The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci USA 1996; 93: 5177-5184

Luan D.D., Korman M.H., Jakubczak J.L., Eickbush T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993; 72: 595-605

Lucas H., Moore G., Murphy G., Flavell R.B. Inverted repeats in the long- terminal repeats of the wheat retrotransposon WIS 2-1A. Mol Biol Evol 1992; 9: 716-728

Ma Z.Q., Roder M., Sorrells M.E. Frequency and sequence characterstics of di-, tri- and tetra-nucleotide microsatellites in wheat. Genome 1996; 39: 123-130

Manninen I., Schulman A. *BARE*-1, a copia- like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 1993; 22: 829-846.

Manninen O., Kalendar R., Robinson J., Schulman A.H. Application of *BARE*-1 retrotransposon markers to map a major resistance gene for net blotch in barley. Mol Gen Genet 2000; 264: 325-334

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

Mellor J., Malim M.H., Gull K., Tuite M.F., McCready S., Dibbayawan T., Kingsman S.M., Kingsman A.J. Reverse transcriptase activity and *Ty* RNA are associated with virus-like particles in yeast. Nature 1985; 318: 583-586

Metzgar D., Bytof J., Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 2000; 10:72-80

Meyers B.C., Tingley S.V., Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 2001; 11: 1660-1676

Miller J.T., Dong F.G., Jackson S.A., Song J.Q., Jiang J. Retrotransposon-related DNA sequences in the centromeres of grass genomes. Genetics 1998; 150: 1615-1623

Moore G., Cheung W., Schwarzacher T., Flavell R. BIS 1, a major component of the cereal genome and a tool for studying genomic organization. Genomics 1991; 19: 469-476

Moore G., Gale M.D., Kurata N., Flavell R.B. Molecular analysis of small grain cereal genomes: current status and prospects. Bio/Technology 1993; 11: 584-589

Moore J.K., Haber J.E. Capture of retrotransposon DNA at the sites of chromosomal double-strand DNA breaks. Nature 1996; 383: 644-645

Morgante M., Oliveri A.M. PCR-amplified microsatellites as markers in plant genetics. Plant Jour 1993; 3: 175-182

Morgante M., Hanafey M., Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 2002; 30: 194-200

Motohashi R., Mochizuki K, Ohtsubo H, Ohtsubo E. Structures and distribution of p-SINE1 members in rice genomes. Theor Appl Genet 1997; 95: 359-368

Muñiz L.M., Cuadrado A., Jouve N., Gonzales J.M. The detection, cloning, and characterisation of WIS 2-1A retrotransposon-like sequences in *Triticum aestivum* and x*Triticosecale* Wittmack and an examination of their evolution in related Triticeae Genome 2001; 44: 978-989

Nadir E., Margalit H., Gallily T, BenSasson S.A. Microsatellite spreadmg in the human genome: Evolutionary mechanisms and structural implications Proc Natl Acad Sci USA 1996; 93: 6470-6475

Nagano H., Wu L., Kawasaki S., Kishima Y., Sano Y. Genomic organization of the 260 kb surrounding the *waxy* locus in a Japonica rice. Genome 1999; 42: 1121-1126

Navarro-Quezada A., Shoen D.J. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. Proc Natl Acad Sci USA 2002, 99: 268-273

Noma K., Ohtsubo E., Ohtsubo H. Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol Gen Genet 1999; 261: 71-79

Nomomura K, Kurata N. Organization of the 1.9 kb repeat unit RCE1 in the centromeric region of rice chromosomes. Mol Gen Genet 1999; 261: 1-10

Nomomura K., Kurata N The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma 2001; 110: 284-291

Panaud O., Chen X., McCouch S R. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome 1995; 38 1170-1176

Panaud O., Chen X, McCouch S R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 1996, 258: 466-478

Panstruga R., Büschges R., Piffanelli P., Schulze-Lefert P. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucl Acids Res 1998; 26: 1056-1062

Pardue M.L., Lowenhaupt K., Rich A., Nordheim A. (dC-dA)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function EMBO Jour 1987; 6:1781-1789

Pardue M.L., Danilevskaya O.N., Lowenhaupt K., Slot F., Traverse K.L. Drosophila telomeres: new views on chromosome evolution. Trends Genet 1996; 12: 48-52

Pedersen C., Linde-Laursen I. Chromosomal location of four minor rDNA loci and a marker microsatellite sequence in barley. Chrom Res 1994; 2: 67-71

Pedersen C, Rasmussen S.K., Linde- Laursen I. Genome and chromosome identification in cultivated barley and related species of the Triticeae (Poaceae) by in situ hybridization with the GAA- satellite sequence. Genome 1996, 39: 93-104

Porceddu A., Albertini E., Baracaccia G., Marconi G, Bertoli F.B., Veronesi F. Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics 2002; 267: 107-114

Powell W., Machray G.C., Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1996; 1: 215-222

Presting G.G., Malysheva L., Fuchs J., Schubert I A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant Jour 1998; 16: 721-728

Ramsay L., Macaulay M., Cardle L., Morgante M., Ivanissevich D.S., Maestri E., Powell W., Waugh R. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant Jour 1999; 17: 415-425

Ramsay L., Macaulay M., Ivanissevich D.S., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvesson S., Morgante M., Massari A., *et al.* A simple sequence repeat-based linkage map of barley. Genetics 2000; 156: 1997-2005

Rashkova S., Athanasiadis A., Pardue M.L. Intracellular targeting of the *Drosophila* telomeric retrotransposons. J Virol 2003; 77: 6376-6384

Roder M.S., Korzun V., Gill B.S., Ganal M.W. The physical mapping of microsatellite markers in wheat. Genome 1998a, 41: 278-283

Roder M.S., Korzun V., Wendehake K., Plaschke J., Tixer M.H., Leroy P., Ganal M.W. A microsatellite map of wheat. Genetics 1998b, 149: 2007-2023

Roder M.S., Plaschke J., Konig S.U., Borner A., Sorells M.E., Tanksley S.D., Ganal M.W. Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 1995; 246: 327-333

Rostocks N., Park Y.J., Ramakrishna W., Ma J., Druka A., Shiloff B.A., SanMiguel P.J., Jiang Z., Brueggeman R., Sandhu D. and others. Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Functional and Integative Genomics 2002; 2: 51-59

Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. Journal of Experimental Medicine 1911; 13: 397-411

Saal B., Wricke G. Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 1999; 42: 964-972.

Sandhu D., Gill K.S. Gene-containing regions of wheat and other grass genomes. Plant Physiol 2002; 128: 803-811

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z. and others. Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

SanMiguel P., Bennetzen J.L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 1998; 82: 37-44

SanMiguel P., Gaut B.S., Tikhoniv A., Nakajima Y., Bennetzen J.L. The paleontology of intergene retrotransposons in maize. Nat Genet 1998; 20: 43-45

Sasaki T. Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Schmidt T. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol 1999; 40: 903-910

Schmidt T., Heslop-Harrison J.S. The physical and genomic hybridization of microsatellites in sugar beet. Proc Natl Acad Sci USA 1996; 93: 8761-8765

Schmidt T., Metzlaff, M. Cloning and characterization of a *Beta vulgaris* satellite DNA family. Gene 1991; 101:247-250

Sharma S., Balyan H.S., Kulwal P.L., Kumar N., Varshney R.K., Prasad M., Gupta P.K. Study of interspecific SSR polymorphism among 14 species from *Triticum-Aegilops* group. Wheat Info Serv 2002; 95:23-28

Shepherd N.S., Schwarz-Sommer Z., Blumberg vel Spalve J., Gupta M., Weinand U., Saedler H. Similarity of the *Cin1* repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 1984; 307: 185-187

Shiba T., Saigo K. Retrovirus-like particles containing RNA homologoues to the transposable element *copia* in *Drosophila melanogaster*. Nature 1983; 302: 119-124

Shimamura M., Yasue H., Ohshima K., Abe H., Kato H., Kishiro T., Goto M., Munechika I., Okada N. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 1997; 388: 666-670

Shinnick T.M., Lerner R.A., Sutcliffe J.G. Nucleotide sequence of Moloney murine leukaemia virus. Nature 1981; 293: 543-548

Shirasu K., Schulman A.H., Lahaye T., Schulze-Lefert P. A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 2000; 10: 908-915

Smilde W.D, Haluskova J., Sasaki T., Graner A. New evidence for the synteny of rice chromosome 1 and barley chromosome 3H from rice expressed sequence tags. Genome 2001; 44: 361-367

Springer P.S., Edwards K.J., Bennetzen J.L. DNA class organization on maize Adh1 yeast artificial chromosomes. Proc Natl Acad Sci USA 1994; 91: 863-867

Suoniemi A., Narvanto A., Schulman A.H. The *BARE-1* retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol Biol 1996a; 31: 295-306

Suoniemi A., Anamthawat-Jónsson K., Arna T., Schulman A.H. Retrotransposon *BARE-1* is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol Biol 1996b; 30: 1321-1329

Suoniemi A., Tanskanen J., Pentikäinen O., Johnson M.S., Schulman A.H. The core domain of retrotransposon integrase in *Hordeum*: Predicted structure and evolution. Mol Biol Evol 1998; 15: 1135-1144

Svitashev S., Bryngelsson T., Vershinin A., Pedersen C., Säll T., von Bothmer R. Phylogenetic analysis of the genus *Hordeum* using repetitive DNA sequences. Theor Appl Genet 1994; 89: 801-810

Takeda S., Sugimoto K., Otsuki H., Hirochika H. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol Biol 1998; 36: 365-376

Tautz D., Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acids Res 1994; 12: 4127-4138

Temin H.M. Nature of the provirus of Rous sarcoma. Natl Cancer Inst Monogr1964; 17: 557-570

Temin H.M., Mizutani S. RNA-directed DNA polymerase in virons of Rous sarcoma virus. Nature 1970; 226: 1211-1213

Temnykh S., Park W.D., Ayres N., Cartinhour S., Hauck N., Lipovich L., Cho Y.G., Ishii T., McCouch S.R. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 2000; 100: 697-712

Temnykh S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11: 1441-1452

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Thiel T., Michalek W., Varshney R.K., Graner A. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 106: 411-422

Thomas C.A. The genetic organization of chromosomes. Annu Rev Genet 1971;5: 237-256. .

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Tikhonov A.P., Lavie L., Tatout C., Bennetzen J.L., Avramova Z., Deragon J.-M. Target sites for SINE integration in Brassica genomes display nuclear matrix binding activity. Chrom Res 2001; 9: 325-337

Trujillo R., Dugan V.L. Radiosensitivity and radiation-induced mutability: an empirical relationship. Radiat Environ Biophys 1975; 12: 253-256

Turcich M.P., Mascarenhas J.P. PREM-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. Sex Plant Reprod 1994; 7: 2-11

Turner B.G., Summers M.F. Structural biology of HIV. J Mol Biol 1999; 285: 1-32

Varagona M.J., Purugganan M., Wessler S.R. Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 1992; 4: 811-820

Varshney R.K. *A study of Some Microsatellites in Some Hexaploid Wheats.* Ph.D. thesis, Ch. Charan Singh University, Meerut, India 2001; pp. 134

Varshney R.K. Kumar A., Balyan H.S., Roy J. K., Prasad M., Gupta P.K. Characterization of microsatellites and development of chromosome specific STMS markers in bread wheat. Plant Mol Biol Rep 2000; 18: 1-12

Varshney R.K., Sharma P.C., Gupta P.K., Balyan H.S., Ramesh B., Roy J.K., Kumar A., Sen A. Low level of polymorphism detected by SSR probes in bread wheat. Plant Breed 1998; 117: 182-184

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Vershinin A.V., Druka A., Alkhimova A.G., Kleinhofs A., Heslop-Harrison J.S. LINE and *gypsy*-like retrotransposons in *Hordeum* species. Plant Mol Biol 2002; 49: 1-14

Vicient C.M., Kalendar R., Anamthawat-Jonsson K., Schulman A.H. Structure, functionality, and evolution of the *BARE*-1 retrotransposon of barley. Genetica 1999a; 107: 53-63

Vicient C.M., Suoniemi A., Anamthawat-Jónsson K., Tanskanen J., Beharav A., Nevo E., Schulman A.H. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum.* Plant Cell 1999b; 11: 1769-1784

Vicient C.M., Kalendar R., Schulman A.H. Envelope-containing retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res 2001a; 11: 2041-2049

Vicient C.M., Jääskeläinen M., Kalendar R., Schulman A.H. Active retrotransposons are a common feature of grass genomes. Plant Physiol 2001b; 125: 1283-1292

Wang Z., Weber J.L., Zhong G., Tanksley S.D. Survey of plant short tandem repeats. Theor Appl Genet 1994; 88: 1-6

Walbot V., Petrov D.A. Gene galaxies in the maize genome. Proc Natl Acad Sci USA 2001; 98: 8163-8164

Waugh R., McLean K., Flavell A.J., Pearce S.R., Kumar A., Thomas B.B.T., Powell W. Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphism (S-SAP). Mol Gen Genet 1997; 253: 687-694

Wei F., Wing R.A., Wise R.P. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 2002; 14: 1903-1917

Weiner A.M. SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 2003; 14: 343-350

Weising K., Nybom H., Wolff K., Meyer W. *DNA Fingerprinting in Plants and Fungi.* CRC Press, Boca Raton, Florida 1995; pp 322

Weising K., Winter P. Huttel B., Kahl G. Microsatellite markers for molecular breeding. Jour Crop Prod 1998; 1: 113-143

Weil C.F., Wessler S.R. The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Biol 1990; 41: 527-552

Wessler S.R. Turned on by stress. Plant retrotransposons. Curr Biol 1996; 6: 959-961

Wessler S.R., Bureau T.E., White S.E. LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. Curr Opin Genet Dev 1995; 5: 814-821

White S.E., Habera L.F., Wessler S.R. Retrotransposons in the flanking regions of normal plant genes: A role for *copia* -like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 1994; 91: 11792-11796

Wicker T., Stein N., Albar L., Feuillet C., Schlagenhauf E., Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant Jour 2001; 26: 307-316

Witte C.P., Le Q.H., Bureau T., Kumar A. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 2001; 98: 13778-13783

Wright D.A., Voytas D.F. *Athila4* of Arabidopsis and *Calypso* of soybean define a lineage of endogenous plant retroviruses. Genome Res 2002; 12: 122-131

Wu K.S., Tanksley S.D. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 1993; 241: 225-235

Xiong Y., Eickbush T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO Jour 1990, 9: 3353-3362

Yamazaki M., Tsugawa H., Miyao A., Yano M., Wu J., Yamamoto S., Matsumoto T., Sasaki T., Hirochika H. The rice retrotransposon *Tois17* prefers low-copy-number sequences as integration targets. Mol Gen Genom 2001; 265: 336-344

Yu G.-X., Wise R.P. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome 2000; 43: 736-749

Yu J., Lu H., Bernardo R. Inconsistency between SSR groupings and genetic backgrounds of white com inbreds. Maydica 2001; 46:133-139

Yu J., Hu S., Wang J., Wong G. K.-S., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *indica*). Science 2002; 296:79-92

Zhong C.X., Marshall J.B., Topp C., Mroczek R., Kato A., Nagaki K., Birchler J.A., Jiang J., Dawe R.K. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 2002; 14: 2825-2836

Zhou S., Ke N., Kim J.M., Voytas D.F. The *Saccharomyces* retrotransposon *Ty5* integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev 1996; 10: 634-645